Human or Computer Assisted Interactive Transcription: Automated Text Recognition, Text Annotation, and Scholarly Edition in the Twenty-First Century

Transcripción humana o asistencia a la transcripción automática interactiva: reconocimiento automático del texto, anotación y edición erudita en el siglo XXI

Transcrição humana ou assistência interativa computadorizada: reconhecimento automático, anotação e edição erudite no século XXI

M. J. CASTRO-BLEDA
J. M. VILAR
S. ESPAÑA
D. LLORENS
A. MARZAL
F. PRAT
F. ZAMORA[1]

Abstract: Computer assisted transcription tools can speed up the initial process of reading and transcribing texts. At the same time, new annotation tools open new ways of accessing the text in its graphical form. The balance and value of each method still needs to be explored. STATE, a complete assisted transcription system for ancient documents, was presented to the audience of the 2013 International Medieval Congress at Leeds. The system offers a multimodal interaction environment to assist humans in transcribing ancient documents: the user can type, write on the screen with a stylus, or utter a word. When one of these actions is used to correct an erroneous word, the system uses this new information to look for other mistakes in the rest of the line. The system is modular, composed of different parts: one part creates projects from a set of images of documents, another part controls an automatic transcription system, and the third part allows the user to interact with the

[1] Departamento de Sistemas Informáticos y Computación, Universitat Politècnica de València, Spain. E-mail: mcastro@dsic.upv.es.

transcriptions and easily correct them as needed. This division of labour allows great flexibility for organising the work in a team of transcribers.

Resumo: Las herramientas de ayuda a la transcripción automática pueden acelerar el proceso inicial de la lectura y transcripción de textos. Al mismo tiempo, las nuevas herramientas de anotación aportan nuevas formas de acceder al texto en su forma original gráfica. Sin embargo, todavía es necesario evaluar las bondades y capacidades de los distintos métodos. STATE, un completo sistema de asistencia a la transcripción de documentos antiguos, se presentó a la audiencia del International Medieval Congress de 2013 celebrado en Leeds. El sistema ofrece un entorno de interacción multimodal para ayudar a las personas en la transcripción de documentos antiguos: el usuario puede teclear, escribir en la pantalla con un lápiz óptico o corregir usando la voz. Cada vez que el usuario cambia de esta forma una palabra, el sistema utiliza la corrección para buscar errores en el resto de la línea. El sistema está dividido en diferentes módulos: uno crea proyectos a partir de un conjunto de imágenes de documentos, otro módulo controla el sistema de transcripción automática, y un tercer módulo permite al usuario interactuar con las transcripciones y corregirlas fácilmente cuando sea necesario. Esta división de las tareas permite una gran flexibilidad para organizar el trabajo de los transcriptores en equipo.

Keywords: Assisted transcription – Ancient documents – Multimodal human/computer interaction – Interactive automatic text recognition.

Palavras-chave: Transcripción asistida – Documentos antiguos – Interacción multimodal persona/computador – Reconocimiento automático de texto interactive.

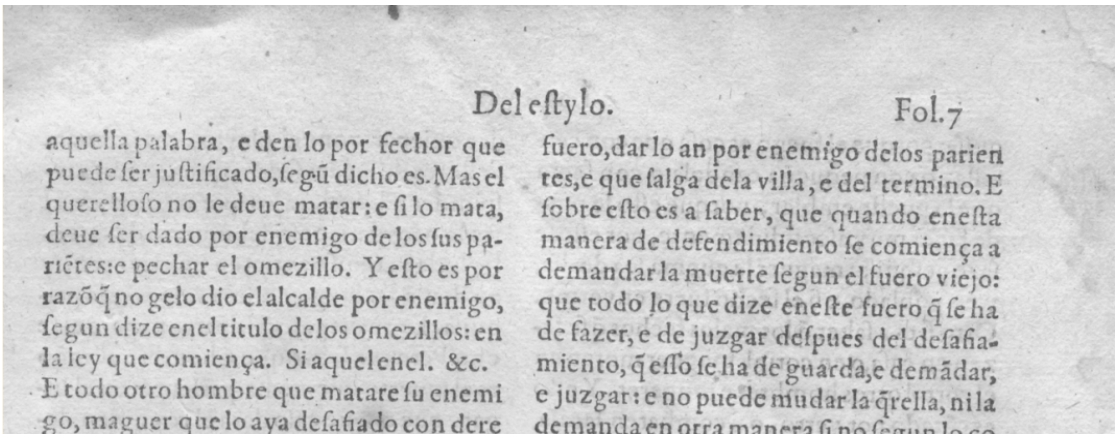
ENVIADO: 16.11.2013
ACEITO: 01.12.2013

***

I. Introduction

The main libraries and document archives are digitising their collections. Most of them are scanning the documents and publishing the resulting images without their corresponding transcriptions. This seriously limits the document exploitation possibilities. The problem is even more serious with respect to ancient documents, both printed (especially from the fifteenth to seventeenth centuries, see Figure 1) and handwritten ones. The commercial Optical Character Recognition Systems offer poor performance on these kinds of documents. When the transcription is necessary, it is performed manually by

experts, which is a very expensive and error-prone task. Having appropriate tools, such specialised recognition engines, would be very helpful to preserve, study and publish the cultural legacy.


Del eſtylo. Fol.7

aquella palabra, e den lo por fechor que puede ſer juſtificado, ſegũ dicho es. Mas el querelloſo no le deue matar: e ſi lo mata, deue ſer dado por enemigo de los ſus parientes: e pechar el omezillo. Y eſto es por razõ q̃ no gelo dio el alcalde por enemigo, ſegun dize en el titulo de los omezillos: en la ley que comiença. Si aquel en el. &c. E todo otro hombre que matare ſu enemigo, maguer que lo aya deſafiado con dere

fuero, dar lo an por enemigo de los parientes, e que ſalga de la villa, e del termino. E ſobre eſto es a ſaber, que quando eneſta manera de defendimiento ſe comiença a demandar la muerte ſegun el fuero viejo: que todo lo que dize eneſte fuero q̃ ſe ha de fazer, e de juzgar deſpues del deſafiamiento, q̃ eſſo ſe ha de guardà, e demãdar, e juzgar: e no puede mudar la q̃rella, ni la demanda en otra manera ſi no ſegun lo co

Figure 1: Example from a Spanish book dated in 1569, printed in ancient font, with archaic lexicon and syntax, lots of abbreviations, and no consistent rule about word separation.

In this context, computer assisted transcription tools can speed up the initial process of reading and transcribing texts. At the same time, new annotation tools open new ways of accessing the text in its graphical form. The balance and value of each method still needs to be explored. STATE, a complete assisted transcription system for ancient documents, was presented in the session titled 'Digital Pleasures V: Automated Text Recognition, Text Annotation, and Scholarly Edition in the 21st Century' last July at the 2013 International Medieval Congress at the University of Leeds.[2] The session focused on how research and development in Computer Science can benefit the Humanities and how enhanced ergonomics may change the landscape of textual scholarship and palaeography (is transcription a computer process assisted by a human or a human activity assisted by a computer?).

Although conventional automatic text recognition systems offer an appropriate performance when dealing with modern documents, their high error rates when recognising unusual fonts make them almost useless for handwritten or ancient documents. Besides, ancient documents are usually

[2] STATE stands for Sistema de Transcripción Asistida de Texto Escrito, which is the Spanish for Written-Text Assisted Transcription System. The current version supports multimodal interaction via mouse, keyboard, stylus, and voice.

affected by many sources of noise (moisture, patches, holes, etc.) that make them harder to transcribe. The STATE system assumes that text recognition performance will frequently be poor on ancient documents and human help will be needed in order to accurately transcribe them; therefore, it is designed as a multimodal, adaptive, extensible system that provides a set of user-centred tools.

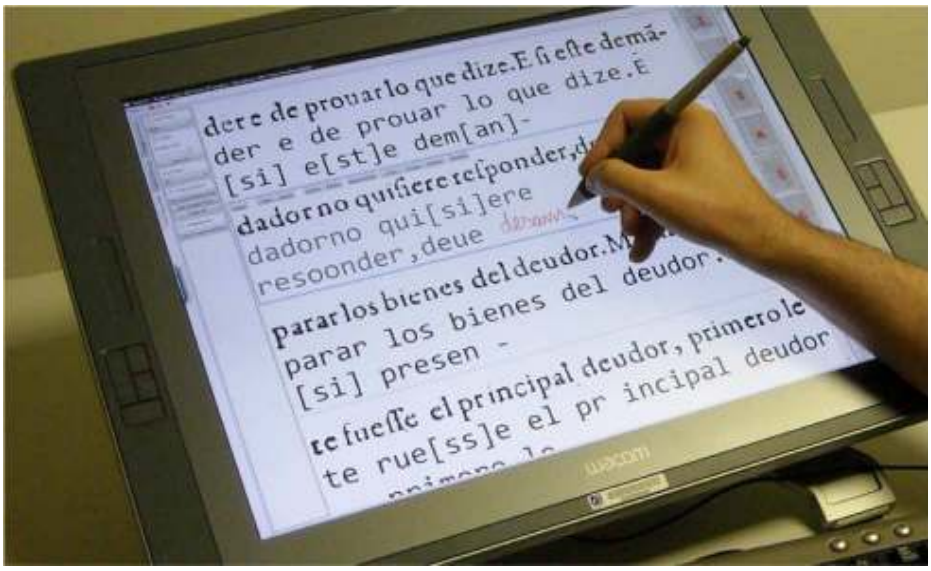


Figure 2: The STATE system.

II. System Overview

STATE has been designed with extensibility and flexibility in mind, which is achieved by providing different applications that are loosely coupled. One application is responsible for creating projects from sets of document pages, another for automatically obtaining transcriptions, and a third one for correcting them. STATE assists the transcription process in a scenario where each document or series of documents to be transcribed can also train a remotely accessible recognition engine used simultaneously by all the users. Each user can interactively manipulate the images (for instance to eliminate stains or increase their contrast) and detect or edit their text layout before transcribing one or more lines of text.

Text transcriptions are shown line by line, together with their original images. When the transcription of a line contains errors, the user can correct them with the keyboard, with an electronic pen, or with the voice, all in a user-

friendly interface. The corrected transcription and its associated image line can then be sent to the recognition engine in order to adapt it to the peculiarities of the task (rare fonts or glyphs, unusual words, ancient syntax, etc.). A privileged user (in the following, the 'administrator') supervises these corrections with a graphical front end and she enhances the engine with those corrected transcriptions. The enhanced recogniser is immediately available to all the users, since it is accessed as a web service.

Summarising, the main applications of the STATE system are:

- A Transcription Assistant that helps to prepare a new project (*project manager*), clean images (*image conditioner*), find page layout to detect blocks of text and lines (*layout manager*), and correct a given transcription line by line (*line transcriber*). Some screenshots of these activities are shown in Figure 3.
- A Text Recognition Engine for ancient printed or handwritten documents.

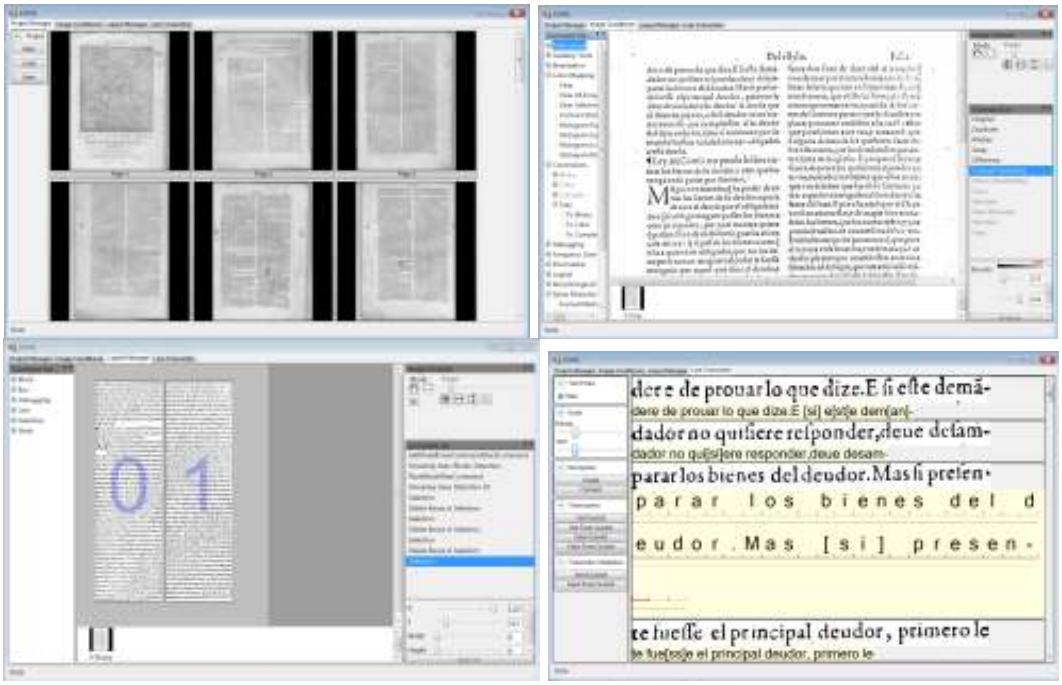


Figure 3: The STATE toolkit: project manager, image conditioner, layout manager, and line transcriber.

The user interface of these applications has been designed to be comfortably used with a stylus, keyboard or mouse. An on-line character recognition engine has been built in order to have a flexible, extensible pen-based text introduction system. This engine is described in detail in 'A Template-based Recognition System for On-line Handwritten Characters'.[3]

III. Interactive Automatic Text Recognition

STATE integrates a particular case of multimodal interaction where one word is corrected by means of keyboard, stylus, or speech at each user-interaction, and where that correction influences on the rest of the line.

The front-end communicates with a back-end recognition engine in order to find the transcription of a line image. This communication is performed by means of a web service described in 'A Flexible System for Document Processing and Text Transcription'.[4] The current version of STATE has two text recognition systems: one based on Nearest Neighbour Search that has proven useful for ancient printed fonts[5] and another recognition system based on Hidden Markov Models hybridised with Artificial Neural Networks (HMM/ANN) or conventional Hidden Markov Models (HMMs) for unconstrained offline handwritten text.[6]

The HMM decoder has been adapted in order to deal with the assisted transcription paradigm: after a sentence has been corrected, there is a consolidated prefix that is considered correct. The idea is that the message

[3] PRAT, F., MARZAL, A., MARTÍN, S., RAMOS-GARIJO, R. and CASTRO, M. J. 'A Template-based Recognition System for On-line Handwritten Characters'. In: *Journal of Information Science and Engineering*, 25: 3, 2009, pp. 779–91.
[4] VILAR, J. M. et al. 'A Flexible System for Document Processing and Text Transcription'. In: *8th Conf. on Technology Transfer (TTIA 2009)*. Sevilla: Sringer, 2009, pp. 467-76.
[5] GORDO, A., LLORENS, D., MARZAL, A., PRAT, F. and VILAR, J. M. 'State: A Multimodal Assisted Text-Transcription System for Ancient Documents'. In: *8th IAPR Workshop on Document Analysis Systems*. Nara: CPS, 2008, pp. 135-42.
[6] ESPAÑA-BOQUERA, S., CASTRO-BLEDA, M. J., GORBE-MOYA, J. and ZAMORA-MARTÍNEZ, F.' Improving Offline Handwritten Text Recognition with Hybrid HMM/ANN Models'. In: *IEEE Trans. Pattern Anal. Mach. Intell.*, 33: 4, 2011, pp. 767–79; ZAMORA-MARTÍNEZ, F., CASTRO-BLEDA, M. J., ESPAÑA-BOQUERA, S. and GORBE-MOYA, J. 'Unconstrained Offline Handwriting Recognition Using Connectionist Character N-grams'. In: *IEEE WCCI IJCNN*. Barcelona: Curran Associates, 2010, pp. 4136-42.

sent to the decoder includes the image of the line together with this consolidated prefix. The consolidated prefix is used by the decoder to force the system to accept this prefix during the first stage of recognition, but the language model probabilities are computed as usual. When available, the confidence measures provided by the ASR are combined with the probability corresponding to the image. In this way, any type of language model such as a combination of several NNLMs as done in [5] can also be.

IV. Conclusions

We have presented the STATE system for assisted transcription of handwritten or printed ancient documents. Great care has been taken in usability aspects and the application can be controlled with mouse, keyboard, stylus, and speech. The design of the interface and its adaptability to new fonts and writing styles allow a great reduction of the transcription effort. The STATE system, thanks to the ergonomics of the final system, incorporates multimodal interfaces that can reduce the effort necessary to correct transcription errors and allows changes in interface (typing, writing with a stylus, or speaking) to reduce fatigue or to follow the preference of the human transcriber, increasing productivity.

Acknowledgments

Work supported by the Spanish Government (TIN2010-18958) and the Generalitat Valenciana (Prometeo/2010/028).